Cache-enabled Small Cell Networks: Modeling and Tradeoffs

Ejder Bastug, Mehdi Bennis, Mérouane Debbah

Cache-enabled Small Cell Networks: Modeling and Tradeoffs

Ejder Baştuğ[◇], Mehdi Bennis[⋆] and Mérouane Debbah[◇],
[◇]Alcatel-Lucent Chair - SUPÉLEC, Gif-sur-Yvette, France
[⋆]Centre for Wireless Communications, University of Oulu, Finland
{ejder.bastug, merouane.debbah}@supelec.fr, bennis@ee.oulu.fi

***Abstract*—We consider the problem of caching in next generation mobile cellular networks where small base stations (SBSs) are able to store their users' content and serve them accordingly. The SBSs are stochastically distributed over the plane and serve their users either from the local cache or internet via limited backhaul, depending on the availability of requested content. We model and characterize the outage probability and average content delivery rate as a function of the signal-to-interference-ratio (SINR), base station intensity, target file bitrate, storage size and file popularity. Our results provide key insights into the problem of cache-enabled small cell networks.**

***Index Terms*—caching, stochastic geometry, limited backhaul, small cell networks, mobile cellular networks**

I. Introduction

Unbearable amount of mobile data demand in the near future [1] is going to introduce new challenges in mobile cellular networks. To meet this huge burden, small cell networks (SCNs) [2], their combination with Wi-Fi [3], heterogeneous networks (HetNets) [4] and many other ideas both from academia and industry are now becoming essential components of future radio access networks. Many European projects such as NewCom# [5] in the 7th Framework Program of the European Commission are focusing on the next generation wireless networks, and many other new projects are expected to take place in the new framework, called Horizon 2020 [6].

Despite this overall view, one particular way of dealing with this crucial problem is to cache the content in the intermediate nodes of the network, resulting in immediate access of the content by users. Content delivery networks (CDNs) such as Akamai [7] and generally speaking information-centric networks (ICNs) [8] are emerging. Combining these concepts in the context of cellular networks is recent [9][10]. Going one step further, it is shown that important gains in terms of user satisfaction and infrastructure cost can be achieved by predicting human behaviour and proactively caching the content in the edge of the network, namely base stations and user terminals [11].

The idea of caching goes back to 60s in the context of operating systems [12]. In the past decades, there has been an extensive literature for web caching schemes (see [13] for example). During the emergence of CDNs, many caching algorithms such as in [14] have appeared. Essentially similar to these solutions but with the difference of shift of paradigms, there have been attempts on solving optimal storage allocation in mobile cellular networks. Due to its non-tractable nature, the vast majority of these recent attempts focus on approximate approaches, or heuristics [15][16][17]. Beside these solutions, information theoretical formulation of the caching problem is studied in [18]. Expected cost of uncoded as well as coded data allocation strategies for stochastically distributed nodes is given in [19]. A game theoretical formulation of the caching problem as a many-to-many game is presented in [20].

In this work, we take a different approach and formulate the caching problem in a scenario where small base stations (SBSs) are stochastically distributed over the plane and have the limited backhaul capacity. Using recent results from [21], we propose a system model and define performance metrics by taking into account system parameters including signal-to-interference-plus-noise ratio (SINR), base station location, target file bitrate, storage size and file popularity distribution. We characterize the performance of our system model for general and specific cases, and validate these results via numerical simulations. To the best of our knowledge, this formulation is the first attempt on coupling caching problem with physical layer (PHY) and enables us to have key design insights for the deployment of cache-enabled SBSs.

The rest of this paper is organized as follows. The system model is described in Section II. Performance metrics and main results related to this system model are provided in Section III. More specific assumptions are made in the same section, yielding much simpler expressions. The results are then validated via numerical simulations in Section IV. We finally conclude and give our possible extensions in Section V.

II. System Model

The cellular network model consists of SBSs modelled as Poisson point process (PPP) Φ with intensity λ. The SBSs are connected to a central scheduler (CS) via wired backhaul links in order to provide broadband connection to their users. The total broadband link capacity given to the CS is finite and fixed. Thus the backhaul link capacity of each SBS, denoted as $C(\lambda)$, is assumed to be a finite decreasing function of λ. The meaning of such an assumption in practice is the following: Deploying more SBSs in a given area, namely increasing λ,

This research has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering), the SHARING project under the Finland grant 128010 and the project BESTCOM.

978-1-4799-5863-4/14/$31.00 ©2014 IEEE

result in the decrement of their backhaul link capacity due to the split of the total broadband link capacity. We will later define $C(\lambda)$ more precisely.

Each SBS has a storage unit with capacity S and employs a strategy by caching users' most popular files. All files have length of L and bitrate requirement of T. We consider that the file popularity distribution of a user is a right continuous and monotonically decreasing probability distribution function (PDF), denoted as $f_{\text{pop}}(f,\gamma)$ where f is a point in the corresponding support of a file and γ is the shape parameter. We further assume that $f_{\text{pop}}(f,\gamma)$ is identical among all users. Therefore, for a given $f_{\text{pop}}(f,\gamma)$, the SBSs store the most popular files before users' request arrival (i.e., off-peak hours during the night).

Using a mobile user terminal, each user is associated with the nearest SBS, resulting in a Voronoi tessellation on the plane. In the downlink, we assume that the tagged SBS transmits with the constant transmit power of $1/\mu$. The standard loss propagation model is used with path loss exponent $\alpha > 2$. The tagged SBS and tagged user experience only Rayleigh fading with mean 1. Hence, the received power of the tagged located r-meters away from its tagged SBS is $hr^{-\alpha}$ where the random variable h follows an Exponential distribution with mean $1/\mu$, denoted as $h \sim \text{Exponential}(\mu)$.

After the user association, the request of a file (or a chunk) is randomly drawn according to the file popularity distribution $f_{\text{pop}}(f,\gamma)$. When this request reaches the SBS via uplink, the user is served immediately in the downlink, by taking the requested file either from internet or from the local cache. If the requested file exists in the local cache, a *cache hit* event occurs. On the contrary, if the requested file is not in the local cache, a *cache miss* event occurs. Our main goal is to analyze the performance of the system during file delivery, hence, the request overhead in the uplink is neglected. An illustration of the network model is shown in Fig. 1.

The performance of our system depends on several factors. For satisfying the quality-of-experience (QoE), the rate of the downlink has to be higher than the file bitrate T, or at least the same. When the cache hit event occurs, assuming that the rate of reading the file from the local disk is sufficiently high, the limiting factor is the rate of the downlink. On the other hand, when the cache miss occurs, the file has to be fetched from the internet via the backhaul. Thus, the rate of backhaul link also contributes to this limitation, besides additional cost of its usage. In the following, we define performance metrics of the system model in details and give our main results. Due to the page limitations, the proofs of these results can be found in the extended version of this work [22].

III. Performance Metrics and Main Results

Our metrics of interest are the *outage probability* and *average delivery rate*. We define these expressions for a downlink cellular network. Without any loss of generality, we refer to the user o as the user located at the origin and called as the *typical* user.

The rate in the downlink depends on the SINR. The SINR of user o at a random distance r from its associated SBS b_o is expressed as:

$$\text{SINR} \triangleq \frac{hr^{-\alpha}}{\sigma^2 + I_r}, \tag{1}$$

where

$$I_r \triangleq \sum_{i \in \Phi / b_o} g_i R_i^{-\alpha}, \tag{2}$$

is the cumulative interference from all the other SBSs except b_o. Supposing that the *success probability* is the probability of the rate exceeding T as well as the probability of the requested file being in the local cache, the outage probability can be defined as the complementary of the success probability as follows:

$$p_{\text{out}}(\lambda, T, \alpha, S) \triangleq 1 - \underbrace{\mathbb{P}[\ln(1+\text{SINR}) > T, f_o \in \Delta_{b_o}]}_{\text{success probability}}, \tag{3}$$

where f_o is the file requested by the typical user, and Δ_{b_o} is the cache of its serving base station b_o. According to our system model and the definition above, we state the following theorem.

Theorem 1. *The outage probability of a typical user from its tagged base station is given by:*

$$p_{\text{out}}(\lambda, T, \alpha, S) = 1 - \pi\lambda \int_0^{\infty} \times \int_0^{S/L} e^{-\pi\lambda v \beta(T,\alpha) - \mu(e^T-1)\sigma^2 v^{\alpha/2}} f_{\text{pop}}(f,\gamma) \mathrm{d}f \mathrm{d}v, \tag{4}$$

where $\beta(T,\alpha)$ is given as:

$$\beta(T,\alpha) = \frac{2\left(\mu(e^T-1)\right)}{\alpha} \times \mathbb{E}_g\left[g^{\frac{2}{\alpha}}\left(\Gamma\left(-\frac{2}{\alpha}, \mu\left(e^T-1\right)g\right) - \Gamma\left(-\frac{2}{\alpha}\right)\right)\right], \tag{5}$$

and Gamma function is defined as $\Gamma(a,x) = \int_x^{\infty} t^{a-1}e^{-t}\mathrm{d}t$ and the incomplete Gamma function is $\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}\mathrm{d}t$.

Beside knowing the probability of a user being in outage, calculating the rate of the delivery is also useful. Delivery rate in our system model is defined as:

$$\tau \triangleq \begin{cases} T, & \text{if } \ln(1+\text{SINR}) > T \text{ and } f_o \in \Delta_{b_o}, \\ C(\lambda), & \text{if } \ln(1+\text{SINR}) > T \text{ and } f_o \notin \Delta_{b_o}, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The main intuition behind this definition is the following. If the downlink rate exceeds the threshold T (namely the bitrate of the requested file) and this file exists in the local cache, the amount of rate given to the user is T, meaning that the tagged SBS provides a minimum required rate. On the other hand, if the downlink rate exceeds T but the file does not exist in the local cache, the file is fetched from the internet via the backhaul, thus the given rate is $C(\lambda)$. We assume that $C(\lambda) < T$ always holds. Given this definition and assumption, we state the following theorem.

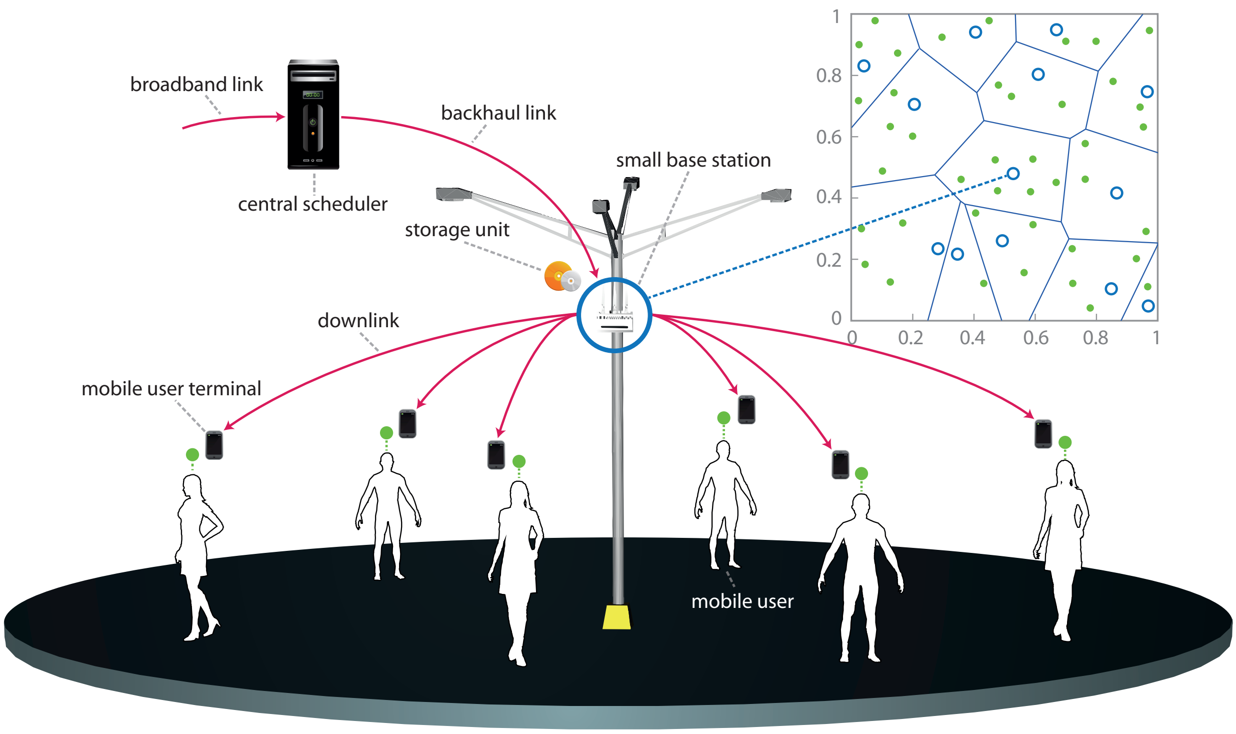

Figure 1: A sketch of the network model. The plot on the top right side is a snapshot of PPP per unit area in which randomly deployed SBSs form a Voronoi tessellation. Communication structure of a SBS, equipped with a storage unit, is highlighted in the main figure.

Theorem 2. *The average delivery rate of a typical user from its tagged base station is given by:*

$$\bar{\tau}(\lambda, T, \alpha, S) = \pi\lambda \int_0^{\infty} e^{-\pi\lambda v \beta(T,\alpha) - \mu(e^T-1)\sigma^2 v^{\alpha/2}} \mathrm{d}v \times \left(C(\lambda) + (T - C(\lambda)) \int_0^{S/L} f_{\mathrm{pop}}(f,\gamma) \mathrm{d}f \right), \quad (7)$$

where $\beta(T,\alpha)$ follows the same definition as in Theorem 1.

Briefly, the results in Theorems 1 and 2 are obtained by averaging over the PPP and fading distributions as well as summation over the popularity distribution to estimate the cache hit ratio. We note that these are general results. After defining $C(\lambda)$, $f_{\mathrm{pop}}(f,\gamma)$ and the distribution of interference more precisely, exact values of the outage probability and average delivery rate can be estimated via numerical integration. In the next section, we derive special cases of these results, which implies much simpler expressions and does not require numerical integration.

A. Special Cases

Assumption 1. *The following assumptions hold in the system model:*

1) *The noise power $\sigma^2 > 0$ and the pathloss component $\alpha = 4$.*
2) *Interference is Rayleigh fading, meaning that $g_i \sim \mathrm{Exponential}(\mu)$.*
3) *Backhaul capacity is defined as*

$$C(\lambda) \triangleq \frac{C_1}{\lambda} + C_2, \quad (8)$$

where $C_1 > 0$ and $C_2 \geq 0$ are some arbitrary design constants such that $C(\lambda) < T$.

4) *The file popularity distribution follows a power law [23] such as*

$$f_{\mathrm{pop}}(f,\gamma) \triangleq \begin{cases} (\gamma-1) f^{-\gamma}, & f \geq 1, \\ 0, & f < 1, \end{cases} \quad (9)$$

where $\gamma > 1$ is the shape parameter.

The assumption $C(\lambda) < T$ comes from the fact that the high-speed fiber-optic backhaul links might be affordable in a scenario where SBSs are densely deployed. Thus, $C(\lambda)$ is assumed to be less than the bitrate of requested files. In practice, the ZipF distribution is used to characterize many real world phenomena (i.e., distribution of files in web-proxies, distribution of word counts in natural languages) and is a discrete power law probability distribution. With the same spirit, we define the power law in continuous domain for our system model. Under the given system model and specific cases made in Assumption 1, we state the following results.

Proposition 1. *The outage probability of a typical user from its tagged base station is given by:*

$$p_{\mathrm{out}}(\lambda, T, 4, S, \gamma) = 1 - \frac{\pi^{\frac{3}{2}}\lambda}{\sqrt{\frac{e^T-1}{\mathrm{SNR}}}} \exp\left(\frac{(\lambda\pi(1+\rho(T,4)))^2}{4(e^T-1)/\mathrm{SNR}}\right) \times Q\left(\frac{\lambda\pi(1+\rho(T,4))}{\sqrt{2(e^T-1)/\mathrm{SNR}}}\right)\left(1 - \left(\frac{L}{L+S}\right)^{\gamma-1}\right), \quad (10)$$

where $\rho(T,4) = \sqrt{e^T - 1}\left(\frac{\pi}{2} - \arctan\left(\frac{1}{\sqrt{e^T-1}}\right)\right)$ *and* $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}\mathrm{d}y$ *is the standard Gaussian tail probability.*

The expression given in Proposition 1 is cumbersome but easy to compute. Indeed, $Q(x)$ given in the expression is a well-known function and can be computed using most modern software packages/libraries without need of integration.

Proposition 2. *The average delivery rate of a typical user from its tagged base station is*

$$\bar{r}(\lambda, T, 4, S, \gamma) = \frac{\pi^{\frac{3}{2}}\lambda}{\sqrt{\frac{e^T-1}{\mathrm{SNR}}}}\exp\left(\frac{(\lambda\pi(1+\rho(T,4)))^2}{4(e^T-1)/\mathrm{SNR}}\right) \times Q\left(\frac{\lambda\pi(1+\rho(T,4))}{\sqrt{2(e^T-1)/\mathrm{SNR}}}\right) \times \left(T + \left(\frac{C_1}{\lambda} + C_2 - T\right)\left(\frac{L}{L+S}\right)^{\gamma-1}\right), \quad (11)$$

where $\rho(T,4)$ *and* $Q(x)$ *are defined as in Proposition 1.*

IV. Numerical Results and Discussion

In this section, we plot the results obtained for outage probability and average delivery rate given in the previous section, and validate them via Monte-Carlo simulations. The simulation setup for each plot is repeated 1000 times and the obtained results are averaged out. We first would like to emphasize that all simulation curves match the theoretical ones. Remember that the results given in the previous section are in the form of continuous variables. Therefore, a slight mismatch observed between theoretical and numerical curves is due to the discretization of these variables. Without loss of generality, we also note that more realistic values can be obtained by making a proper SINR gap approximation and taking into account the total wireless bandwidth. On the other hand, target file bit rate and average delivery rate are in the units of nats/sec/Hz (1 bit = ln(2) $= 0.693$ nats). Moreover, storage size and file lengths are in the units of nats.

One important design parameter in our scenario is the storage size of SBSs. In order to see its impact, the evolution of the outage probability and average delivery rate with respect to the storage size is shown in Fig. 2. Additionally, for each curve, the results are shown for different values of target file bit rate. Clearly, by increasing the storage size, the outage probability reduces and its daughter metric average delivery increases at the same time. Such a behaviour obtained from theoretical and simulation curves confirms our intuition.

The impact of number of the base stations on the outage probability is depicted in Fig. 3. From the figure, it can be shown that increasing the base station density results in a decrement of outage probability. Moreover, further decrease in outage probability can be achieved by increasing the storage size of SBSs.

In our setup, yet another design parameter is the target file bit rate T. To show its importance, the change of outage probability with respect to the target file bit rate is given in Fig. 4. The curves are shown for different values of storage size. Obviously, the outage probability increases with the increment of target bitrate. This increment in outage probability can be reduced by increasing the storage size of SBSs. However, as T

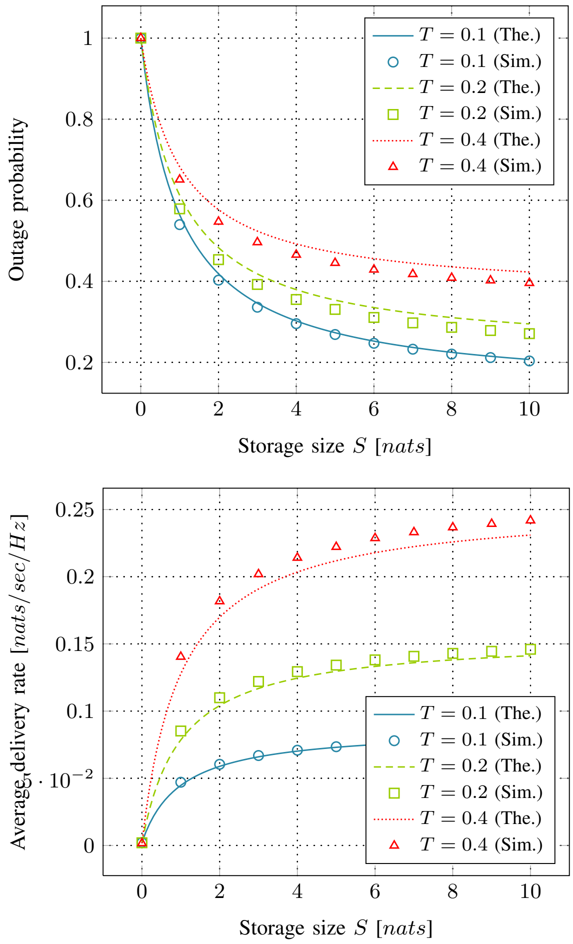

Figure 2: Evolution of outage and average delivery rate with respect to storage size. SNR $= 10$ dB, $\lambda = 0.2$, $\gamma = 2$, $L = 1$, $\alpha = 4$, $C_1 = 0.0005$, $C_2 = 0$.

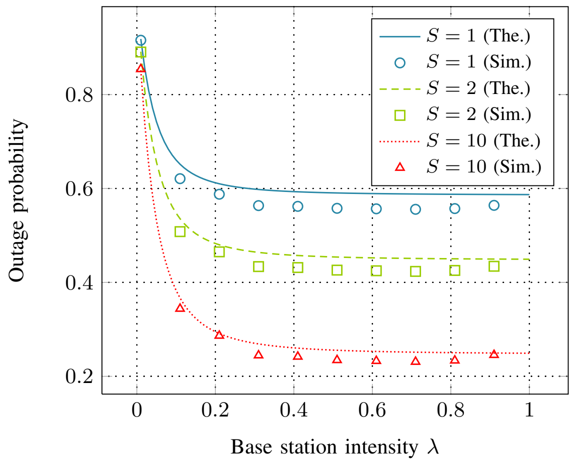

Figure 3: Evolution of outage with respect to base station intensity. SNR $= 10$ dB, $T = 0.2$, $\gamma = 2$, $L = 1$, $\alpha = 4$, $C_1 = 0.0005$, $C_2 = 0$.

increases, the impact of storage size on the outage probability reduces.

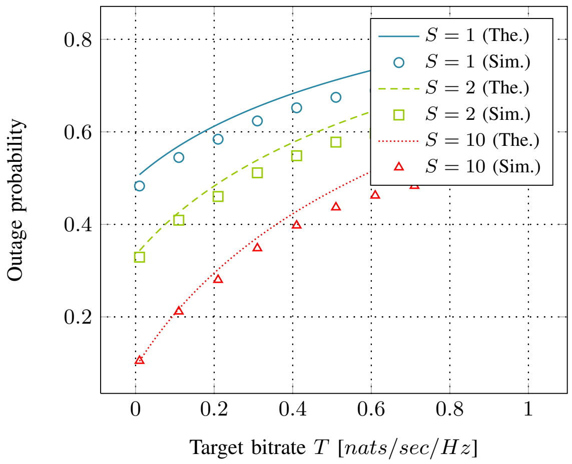

Figure 4: Evolution of outage with respect to target bitrate. SNR $= 10$ dB, $\lambda = 0.2$, $\gamma = 2$, $L = 1$, $\alpha = 4$, $C_1 = 0.0005$, $C_2 = 0$.

Another crucial parameter in our scenario is the shape of the popularity distribution, characterized by the parameter γ. The impact of this shape is shown in Fig. 5 for different storage sizes. In general, as γ increases, some portion of files becomes more popular than others. Thus an increase of γ results in a decrement of outage probability. Moreover, in very low and high values of γ, we observe that the storage size has a small impact on the outage probability.

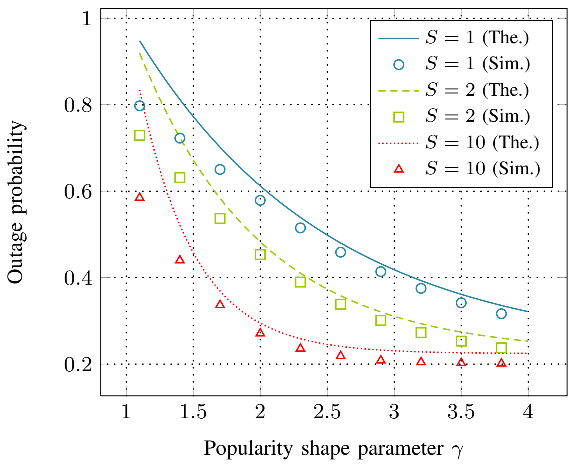

Figure 5: Evolution of outage with respect to popularity shape parameter. SNR $= 10$ dB, $\lambda = 0.2$, $\gamma = 2$, $L = 1$, $\alpha = 4$, $C_1 = 0.0005$, $C_2 = 0$.

V. Conclusions

We have studied the problem of caching where cache-enabled SBSs are stochastically distributed and have limited backhaul capacity. We derived expressions for outage and average delivery rate, then validated our results via numerical simulations. The results for each parameter of interest showed that several gains are possible by employing storage units in SBSs. With this in mind, telecom operators can either deploy more base station or increase total storage size to guarantee a certain outage probability for QoE. For this deployment problem, an interesting future work would be investigating the trade-off between density of base stations and storage size.

References

[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018," *White Paper, [Online] http://goo.gl/l77HAJ*, 2014.

[2] J. Hoydis, M. Kobayashi, and M. Debbah, "Green Small-Cell Networks," *IEEE Vehicular Technology Magazine*, vol. 6(1), pp. 37–43, 2011.

[3] M. Bennis, M. Simsek, W. Saad, S. Valentin, M. Debbah, and A. Czylwik, "When Cellular Meets WiFi in Wireless Small Cell Networks," *IEEE Communication Magazine, Special Issue in HetNets*, June 2013.

[4] J. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 136–144, 2013.

[5] Newcom#, "Network of Excellence in Wireless Communications," *[Online] http://www.newcom-project.eu*, 2014.

[6] Horizon 2020, "The EU Framework Programme for Research and Innovation," *[Online] http://ec.europa.eu/programmes/horizon2020*, 2014.

[7] E. Nygren, R. K. Sitaraman, and J. Sun, "The akamai network: a platform for high-performance internet applications," *ACM SIGOPS Operating Systems Review*, vol. 44, no. 3, pp. 2–19, 2010.

[8] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," *Communications Magazine, IEEE*, vol. 50, no. 7, pp. 26–36, 2012.

[9] S. Spagna, M. Liebsch, R. Baldessari, S. Niccolini, S. Schmid, R. Garroppo, K. Ozawa, and J. Awano, "Design principles of an operator-owned highly distributed content delivery network," *Communications Magazine, IEEE*, vol. 51, no. 4, pp. 132–140, April 2013.

[10] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: exploiting content caching and delivery techniques for 5g systems," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 131–139, February 2014.

[11] E. Baştuğ, M. Bennis, and M. Debbah, "Living on the Edge: The role of Proactive Caching in 5G Wireless Networks," *IEEE Communications Magazine*, In Press, August, 2014.

[12] L. A. Belady, "A study of replacement algorithms for a virtual-storage computer," *IBM Syst. J.*, vol. 5, no. 2, p. 78–101, 1966.

[13] J. Wang, "A survey of web caching schemes for the internet," *ACM SIGCOMM Computer Communication Review*, vol. 29, no. 5, pp. 36–46, 1999.

[14] S. Borst, V. Gupta, and A. Walid, "Distributed caching algorithms for content distribution networks," in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–9.

[15] E. Baştuğ, J.-L. Guénégo, and M. Debbah, "Proactive small cell networks," in *20th International Conference on Telecommunications (ICT)*, Casablanca, Morocco, 05/2013 2013.

[16] K. Poularakis, G. Iosifidis, V. Sourlas, and L. Tassiulas, "Multicast-aware caching for small cell networks," in *IEEE Wireless Communications and Networking Conference (WCNC 2014)*, Istanbul, Turkey, April 2014.

[17] P. Blasco and D. Gunduz, "Learning-based optimization of cache content in a small cell base station," in *IEEE International Conference on Communications (ICC 2014)*, Sydney, Australia, June 2014.

[18] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 1077–1081.

[19] E. Altman, K. Avrachenkov, and J. Goseling, "Coding for caches in the plane," *arXiv preprint arXiv:1309.0604*, 2013.

[20] K. Hamidouche, W. Saad, and M. Debbah, "Many-to-many matching games for proactive social-caching in wireless small cell networks," in *WNC3 workshop in conjuction with WiOpt*, Hammamet, Tunisia, 2014.

[21] J. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122–3134, 2011.

[22] E. Baştuğ, M. Bennis, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," *In Preperation*, 2014.

[23] M. E. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.